

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12Q 1/68	A1	(11) International Publication Number: WO 98/44152 (43) International Publication Date: 8 October 1998 (08.10.98)
(21) International Application Number: PCT/GB98/00963 (22) International Filing Date: 1 April 1998 (01.04.98) (30) Priority Data: 9706529.6 1 April 1997 (01.04.97) GB 9713236.9 23 June 1997 (23.06.97) GB (71) Applicant (for all designated States except US): GLAXO GROUP LIMITED [GB/GB]; Glaxo Wellcome House, Berkeley Avenue, Greenford, Middlesex UB6 0NN (GB). (72) Inventors; and (75) Inventors/Applicants (for US only): KAWASHIMA, Eric, H. [US/CH]; Seroxo Pharmaceutical Research Institute S.A., 14, chemin des Aulx, CH-1228 Plan Les Ouates (CH). FARINELLI, Laurent [CH/CH]; 15, chemin des Aulx, CH-1256 Troinex (CH). MAYER, Pascal [FR/CH]; Seroxo Pharmaceutical Research Institute S.A., 14, chemin des Aulx, CH-1228 Plan Les Ouates (CH). (74) Agent: TEUTEN, Andrew, J.; Glaxo Wellcome plc, Glaxo Wellcome House, Berkeley Avenue, Greenford, Middlesex UB6 0NN (GB).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: METHOD OF NUCLEIC ACID SEQUENCING (57) Abstract Different nucleic acid molecules present at different locations can be sequenced in parallel. Primers that are annealed to the nucleic acid molecules can be provided. Each location can then be provided with a nucleic acid polymerase and a nucleotide. It can then be determined whether or not the nucleotide has been used in primer extension and the process can be repeated. As an alternative to using primers, a nick in a double stranded nucleic acid molecule can provide a 3'-OH group for chain extension.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Larvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

METHOD OF NUCLEIC ACID SEQUENCING

The present invention relates to nucleic acid sequencing.

5 Nucleic acid sequencing methods have been known for many years. Two of the best known methods are the Sanger "dideoxy" method and the Maxam-Gilbert method.

The Sanger dideoxy method relies upon dideoxyribonucleoside triphosphates
10 being used as chain terminators. A DNA polymerase is used to extend a primer in order to synthesise a DNA sequence that is complementary to at least part the sequence of a target molecule. The primer extension reaction is performed in four different batches, each batch containing four radioactive
15 deoxyribonucleoside triphosphates and a dideoxy analogue of one of the deoxyribonucleoside triphosphates. The dideoxy analogue is an analogue of a different deoxyribonucleoside triphosphate in each batch. Once an analogue has been incorporated then no further primer extension can occur and the length of the strand formed is fixed. The strands of different length formed in each of the four batches are then size-fractionated in each of four different lanes of a gel
20 (corresponding to the four different batches). This allows the sequence to be read from an autoradiograph.

In contrast to the Sanger dideoxy method, the Maxam-Gilbert method relies upon the chemical cleavage of labelled DNA. Here several identical DNA
25 strands are each labelled at one end with a radioactive label. The strands are then split into four different groups. These groups are subjected to particular reaction conditions, which are different for each group. The reaction conditions

are chosen to cause an average of about one chemical cleavage reaction per strand at a location where a particular base is present. The cleavage products of the four groups are then size fractionated on four different lanes of a gel (corresponding to cleavage at each of four different bases). By using
5 autoradiography to read the gel the order in which the bases appear in the original DNA sequence can be determined.

Although the methods discussed above have been used extensively, they rely upon time-consuming sequencing from gels and are therefore not for the
10 parallel sequencing of many different samples. Other methods, which do not require the use of gels, have therefore been designed for parallel sequencing.

WO 96/12039 (Lynx Therapeutics, Inc.) discloses a method for determining the nucleotide sequence of a target polynucleotide. The method comprising the
15 steps of generating from the target polynucleotide a plurality of fragments that cover the target polynucleotide; attaching an oligonucleotide tag from a repertoire of tags to each fragment of the plurality such that substantially all the same fragments have the same oligonucleotide tag attached and substantially all different fragments have different oligonucleotide tags attached, the
20 oligonucleotide tags being oligonucleotides of natural nucleotide monomers having a length in the range of from 10 to 20 nucleotides or oligonucleotides of natural nucleotide monomers comprising a plurality of subunits such that each subunit of the plurality consists of an oligonucleotide of natural nucleotide monomers having a length from three to six nucleotides, the subunits being
25 selected from a minimally cross-hybridising set; sorting the fragments by specifically hybridising the oligonucleotide tags with their respective tag complements; determining the nucleotide sequence of a portion of each of the

fragments of the plurality; and determining the nucleotide sequence of the target polynucleotide by collating the sequences of the fragments.

In one embodiment of the method disclosed in WO 96/12039, the step of
5 determining the nucleotide sequence of the fragments is carried out simultaneously for a plurality of fragments by a single base sequencing method. A suitable single base sequencing method is said to be the method disclosed in international patent application PCT/US95/03678 (published as WO95/27080 -
Lynx Therapeutics, Inc.). This method comprises the steps of: ligating a probe
10 to an end of a polynucleotide, the probe having a nuclease recognition site; identifying one or more nucleotides at the end of the polynucleotide; and cleaving the polynucleotide with a nuclease recognising the nuclease recognition site of the probe such that the polynucleotide is shortened by one or more nucleotides.

15 WO96/12014 (Lynx Therapeutics Inc.) also describes the molecular tagging system that is described in WO96/12039. However here various applications in addition to sequencing are discussed.

20 WO93/21340 (The Medical Research Council) discloses a sequencing method. This method comprises the steps of : forming a single-stranded template comprising the nucleic acid to be sequenced; hybridising a primer to the template to form a template/primer complex; extending the primer by the addition of a single labelled nucleotide; determining the type of the labelled
25 nucleotide added onto the primer; removing or neutralising the label; repeating the preceding two steps sequentially and recording the order of incorporation of

labelled nucleotides. The method is said to be suitable for use in an automated procedure whereby a large number of templates are sequenced simultaneously.

DE-A-4141178 (EMBL) discloses a method similar to that disclosed in
5 WO93/21340. This- is also said to be suitable for the simultaneous sequencing of several nucleic acids.

The method is for sequencing an at least partially single-stranded nucleic acid used as a template and uses a suitable polymerase together with four nucleotides
10 in labelled form. Starting from a primer or a double-stranded nucleic acid section, a complementary strand is synthesised in a step-by step manner using the labelled nucleotides. The method includes the steps of incubating a nucleic acid which is provided with a primer or which is partially double-stranded with an incubation mixture consisting of a polymerase, one of the four labelled
15 nucleotides and other substances needed for chain polymerisation so that a labelled nucleotide is incorporated into a growing complementary strand in the event that the next available nucleotide on the template-strand is complementary to the labelled nucleotide; separating the nucleic acid strand which may have been extended by one nucleotide in the previous step and
20 carrying out one or more washing steps in known manner; determining whether a nucleotide has been incorporated by virtue of its label; and, if a nucleotide has successfully been incorporated, removing the label. The preceding steps are cyclically executed for all four nucleotides.

25 From the foregoing description it will be appreciated that several methods for parallel sequencing are known. However, although known methods for parallel sequencing have greatly increased the number of samples that can be sequenced

in a given time, there is a need to provide an even greater throughput of samples. Many groups are attempting to sequence very large numbers of nucleic acid molecule samples and even complete genomes (as in the human genome sequencing project, for example). However they are limited by the efficiency of the sequencing techniques available. Furthermore most existing methods of parallel sequencing are costly, often requiring the use of large numbers of reagents and/or requiring many process steps. The presence of many different reagents and the use of many process steps is disadvantageous in that it can lead to sequencing errors.

10

The present invention aims to alleviate one or more of the disadvantages of known parallel sequencing methods.

According to the present invention there is provided a method for sequencing nucleic acid molecules, comprising the steps of:

15

- a) providing at a first location a plurality of single stranded nucleic acid molecules that have the same sequences as one another and that are hybridised to primers in a manner to allow primer extension in the presence of nucleotides and a nucleic acid polymerase;
- b) providing at a second location, which is different from the first location, a plurality of single stranded nucleic acid molecules that have the same sequences as one another, but that have different sequences from the sequences of the single stranded nucleic acid molecules at the first location, and that are also hybridised to primers in a manner to allow primer extension in the presence of nucleotides and a nucleic acid polymerase;

25

c) providing each location with a nucleic acid polymerase and a given labelled nucleotide under conditions that allow extension of the primers if a complementary base or if a plurality of such bases is present at the appropriate position in the single stranded nucleic acid molecules;

d) detecting whether or not said labelled nucleotide has been used for primer extension at each location by determining whether or not the label present on said nucleotide has been incorporated into extended primers;

e) repeating steps c) and d) one or more times so that extended primers comprising a plurality of labels are provided.

By providing different template molecules at different locations, the present invention allows the sequencing of different nucleic acid molecules to occur in parallel.

The present invention allows labelled nucleotides to be incorporated in a stepwise manner in a nucleic acid molecule via primer extension. However the presence of one labelled nucleotide in the molecule does not prevent other labelled nucleotides being detected (even if adjacent labelled nucleotides are the same). Thus, unlike many prior art parallel sequencing methods, such as those disclosed in WO93/21340 and DE-A-4141178, there is no need to remove a label from a polynucleotide chain before a further labelled nucleotide is added (although in some embodiments it may be desired to remove labels periodically). A plurality of labels can therefore be incorporated into a nucleic acid molecule via primer extension and can be detected *in situ*.

Furthermore, as discussed in greater detail later on, in one embodiment the present invention is particularly advantageous in sequencing nucleic acid strands comprising one or more stretches of contiguous identical bases (e.g. polyA, polyT, polyC, polyU or polyG stretches). This is because the number of steps required for sequencing such strands can be significantly reduced.

The present invention is also advantageous over prior art methods that rely upon the use of large numbers of different oligonucleotide tags in determining nucleotide sequences, such as the methods disclosed in WO96/12014 and WO96/12039. These methods are generally time-consuming and require specific tags to be provided. They may also use restriction enzyme digestion or other means to remove tags or parts thereof following a detection step. The present invention does not suffer from these drawbacks.

A further advantage of the present invention over various prior art methods is that it allows high sensitivity because labels can be directly incorporated into a plurality of identical nucleic acid molecules that are present at a single location. As described later, the incorporation of labelled nucleotides (e.g. fluorescence-labelled nucleotides) and detection by a sensitive cooled CCD camera can allow a sharp signal to be produced and detected *in situ*. The CCD camera is preferably mounted on an inverted microscope.

The single stranded molecules that are hybridised to primers in steps a) and b) of the method of the present invention are referred to herein as "templates". These may be DNA or RNA molecules. They may comprise naturally occurring and / or non-naturally occurring nucleotides.

Preferably the templates are immobilised on a surface - i.e. they are maintained in position on a surface by covalent or non-covalent interactions for sufficient time to allow the present invention to be performed. They may be directly or indirectly bound to the surface.

Sequencing may be full or may be partial. The sequence of a complementary strand to the template or of a part thereof can be obtained initially. However this sequence can be converted (using base-pairing rules) to provide the sequence of the template or of a part thereof. This conversion can be done via a computer or via a person. It can be done after each step of primer extension or at a later stage.

Desirably, the method of the present invention includes a step of removing labelled nucleotides from the first and second locations if they are not incorporated by primer extension (e.g. via a washing step, as will be described later). By immobilising templates, unincorporated labelled nucleotides can be removed from the locations of the templates whilst the templates are maintained in position.

Any suitable surface may be provided for immobilising nucleic acid molecules. The term "support" is used herein to describe a material possessing such a surface. The support is preferably a solid material or a semi-solid material (e.g. a gel or a lipidic monolayer / bilayer).

Various protocols are known for binding nucleic acids to a surface. For example they can be covalently bound to the surface of a glass or polymer

- support (see e.g., Joos, B., Kuster, H., and Cone, R. (1997), entitled "Covalent attachment of hybridizable oligonucleotides to glass supports", *Analytical Biochemistry* 247, 96-101; Oroskar AA, Rasmussen SE, Rasmussen HN, Rasmussen SR, Sullivan BM, Johansson A. (1996), entitled "Detection of immobilized amplicons by Elisa-like techniques", *Clinical Chemistry* 42(9):1547-1555; Khandjian EW, entitled "UV crosslinking of RNA to nylon membrane enhances hybridization signals", *Molecular Biology Reports* 11(2):107-15, 1986.)
- 10 Nucleic acid molecules may also be immobilised to a support via non-covalent binding. The non-covalent binding may be achieved, for example, by interaction between biotinylated-nucleic acids and a (strept)avidin coated surface or by the anchoring of a hydrophobic chain into a lipidic monolayer or bilayer.
- 15 Indeed any interaction between a surface and a nucleic acid molecule which allows the various steps of the method to be performed may be utilised for binding. For example template molecules may be bound to a surface merely by base pairing with primers already bound to a surface.
- 20 In preferred embodiments, a single surface is provided on which template molecules are immobilised. In less preferred embodiments, discrete units each having their own surface may be provided. For example, spheroids may be provided (e.g. glass beads/beads formed of another polymeric material, etc.).
- 25 Alternatively other three-dimensional units may be provided.

The discrete units may possess magnetic properties or other properties to facilitate separation of the units from other components, e.g. during washing steps. (Magnetic beads can be obtained from Dynabeads™ M-280, Dynal A.S. Oslo, Norway. They can be used as described by Hultman, T., Bergh, S., Moks, T., Uhlen, M. Bidirectional solid-phase sequencing of *in vitro*-amplified plasmid DNA. BioTechniques 10:84-93, 1991.)

Different units may even have different properties in order to facilitate separation into different categories of unit (e.g. primers of one type may be bound to units having a given property and primers of another type may be bound to units having a different property or at least not having the given property).

Desirably, a planar surface is provided (e.g. by a support which is a membrane or is a glass slide).

It is preferred to provide high densities of immobilised template molecules and/or of primers. Desirably high densities are provided in small areas. This can be achieved using robotic systems. Such systems may rely on piezoelectric delivery systems or other delivery systems in order to deliver small quantities (e.g. nanolitres or even picolitres) of material to small areas. The small areas may be in the form of arrays (that may be regular or irregular). By controlling the operation of a delivery device, different types of array may be provided.

In less preferred embodiments arraying can be performed after binding of template and/or primer molecules to a surface. In such a case the molecules can be immobilised on discrete units of support (e.g. as discussed above), which are

subsequently arrayed. Each unit may therefore have a surface on which immobilised nucleic acids are present.

5 Desirably, a large number of copies of the same initial template molecule is provided at a given location. For example, over 10,000 such molecules may be provided.

10 A location may be in the form of a single distinct area (that is preferably planar). Preferably the areas are from 100 nm to 2 cm, more preferably from 500 nm to 5 mm in length, when measured across the largest dimension. In the case of areas defined by a generally circular perimeter this measurement will be the diameter of the circle.

15 Less preferably, a location may comprise the contents of a container or a part thereof (e.g. template-bound beads in a tube or in a microplate well).

A plurality (two or more) of locations having one or more of the characteristics discussed in the foregoing paragraphs will be generally provided.

20 Where a large number of template molecules are provided at the locations, this can enhance the reliability of sequence data provided. This is because the probability of a false reading being provided due to the presence of contaminants can be minimised.

25 In preferred embodiments, the sequencing may be performed on a small number of template molecules provided that sensitive detection techniques are used.

Sensitive detection techniques allowing for detection of fluorescently-labelled molecules can, for example, utilise:

- 5 a) A cooled Charge Coupled Device (CCD) camera (e.g. Princeton Instruments),
- b) Confocal microscopy (e.g. Carl Zeiss Jena),
- or
- c) High-speed DNA sequencing: an approach based upon fluorescence detection of single molecules. (Journal of Biomolecular Structure & Dynamics
- 10 7: 301-309, Jett, J.H., Keller, R.A., Martin, J.C., Marrone, B.L., Moyzis, R.K., Ratliff, R.L., Seitzinger, N.K., Brooks Shera, E., Stewart, C.C. (1989)).

Template molecules provided at different locations may be nucleic acid molecules from different sources. These may or may not have one or more

15 identical/homologous sequences along their lengths. For example, different template molecules may have identical/homologous sequences over all/part of their lengths when similar samples from different organisms are provided (e.g. similar genetic regions from related organisms). Alternatively, different samples from the same organism (e.g. samples from different types of cell,

20 tissue or organ) may be provided.

One of the main advantages of the present invention is that sequences, whether full or partial, can be read simultaneously and efficiently from a plurality of different locations. (The present invention is not limited to being applied at

25 only the first and second locations discussed previously.) For example, over 10, over 100, over 1000, or even over 1,000,000 different locations may be

provided. Thus many different sequences can be determined in a relatively short time period.

5 If desired, each type of template molecule can be provided at a plurality of different locations (preferably in amplified form at each location), allowing for redundancy of the sequence data generated. This redundancy allows controls to be provided to provide checks on the accuracy of sequence data generated.

10 Oligonucleotides are preferred as primers suitable for use in the present invention. These are nucleic acid molecules that are typically 6 to 60, e.g. 15 to 25 nucleotides long. They may comprise naturally and / or non-naturally occurring nucleotides. Other molecules, e.g. longer nucleic acid strands may alternatively be used as primers, if desired.

15 Primers can be annealed (hybridised) to template molecules. Primers which remain in solution or which do not anneal specifically to the template are preferably removed after annealing. Preferred annealing conditions (temperature and buffer composition) prevent non-specific hybridisation. These may be stringent conditions. Such conditions would typically be annealing
20 temperatures close to a primer's T_m (melting temperature) at a given salt concentration (e.g. 50 nM primer in 200 mM NaCl buffer at 55°C for a 20-mer oligonucleotide with 50% GC content).

(Stringent conditions for a given system can be determined by a skilled person.
25 They will depend on the base composition, GC content, the length of the primer used and the salt concentration. For a 20 base oligonucleotide of 50% GC,

calculated average annealing temperature is 55-60°C, but in practice may vary between 35 to 70°C).

5 Primers may be immobilised by being directly linked to a surface by the techniques discussed *supra* for template molecules. Alternatively, they may be indirectly linked to the surface, e.g. by virtue of being annealed to a template molecule that is itself bound to the surface.

10 In any event the template molecule will comprise a portion that hybridises with the primer (preferably under "stringent" conditions). This portion can be added to a given molecule (even if of a totally/partially unknown sequence) using techniques known to those skilled in the art to provide a template. For example it can be synthesised artificially and can be linked via a ligase to the molecule of totally/partially unknown sequence. This can result in a single or double
15 stranded molecule. If a double stranded molecule is produced, heating can be used to separate annealed strands to provide a single stranded template, which can then be annealed to a primer.

20 In addition to the ways of providing a template annealed to a primer discussed above, a template annealed to a primer can also be provided by starting with a double stranded molecule and then removing part of one of the strands. This will leave a part of one strand (the primer) annealed to a longer strand. Here the longer strand will comprise a single stranded portion to be sequenced.

25 The step of removing part of one of the strands can be achieved for example by using one or more nucleases. One way of doing this is by using limited digestion with a 3'→5' exonuclease to leave a short strand with a free 3' OH

group (the primer) hybridised to a larger strand (the template). (The longer strand can be capped at its 3' end to prevent it being digested by the exonuclease.) Another way is to introduce a nick into one strand with an endonuclease and then to remove part of that strand with a 5'→3' exonuclease.

5 This will leave a free 3' OH group on a short strand (the primer) hybridised to a longer strand (the template). (The longer strand can be capped at its 5' end to prevent it being digested by the exonuclease.)

By whatever means the primer annealed to the template is provided, as

10 discussed previously, it is preferred that the primer or the template molecule is immobilised. Immobilisation can be performed either before or after providing the primer annealed to the template.

Once a template annealed to a primer is provided, primer extension can be

15 performed. RNA or DNA polymerases can be used. DNA polymerases are however the enzymes of choice for preferred embodiments. Several of these are commercially available. Polymerases which lack 3' → 5' exonuclease activity can be used, such as T7 DNA polymerase or the small (Klenow) fragment of DNA polymerase I may be used [e.g. the modified T7 DNA

20 polymerase Sequenase™ 2.0 (Amersham) or Klenow fragment (3'→5' exo-, New England Biolabs)]. However it is not essential to use such polymerases. Indeed, where it is desired that the polymerases have proof-reading activity polymerases lacking 3' → 5' exonuclease activity would not be used.

25 Certain applications may require the use of thermostable polymerases such as ThermoSequenase™ (Amersham) or Taquenase™ (ScienTech, St Louis, MO).

Any nucleotides may be used for primer extension reactions (whether naturally occurring or non-naturally occurring). Preferred nucleotides are deoxyribonucleotides; dATP, dTTP, dGTP and dCTP (although for some applications the dTTP analogue dUTP is preferred) or ribonucleotides ATP, UTP, GTP and CTP; at least some of which are provided in labelled form.

The use of labelled nucleotides during primer extension facilitates detection. (The term "label" is used in its broad sense to indicate any moiety that can be identified using an appropriate detection system. Preferably the label is not present in naturally occurring nucleotides.)

Ideally, labels are non-radioactive, such as fluorophores which allow efficient detection of primer extension.

In some applications, considering that a large number of copies of each template molecules can be immobilised at each location, one can envisage the use of a combination of labelled and non-labelled nucleotides. In this case, even if a small proportion of the incorporated nucleotides are labelled (e.g. fluorescence-labelled), the number of labels at each location can be sufficient to be detected by a detection device. For example the ratio of labelled:non labelled nucleotides may be chosen so that labelled nucleotides are used in primer extension less than 50%, less than 20% or less than 10% of the time.

Thus in one embodiment of the present invention there is provided a method for sequencing nucleic acid molecules, comprising the steps of:

- a) providing at a first location a plurality of single stranded nucleic acid molecules that have the same sequences as one another and that are

hybridised to primers in a manner to allow primer extension in the presence of nucleotides and a nucleic acid polymerase;

- 5 b) providing at a second location, which is different from the first location, a plurality of single stranded nucleic acid molecules that have the same sequences as one another, but that have different sequences from the sequences of the single stranded nucleic acid molecules at the first location, and that are also hybridised to primers in a manner to allow primer extension in the presence of nucleotides and a nucleic acid polymerase;
- 10
- c) providing each location with a nucleic acid polymerase and a given nucleotide in labelled and unlabelled form under conditions that allow extension of the primers if a complementary base or if a plurality of such bases is present at the appropriate position in the single stranded nucleic acid molecules;
- 15
- d) detecting whether or not said labelled nucleotide has been used for primer extension at each location by determining whether or not the label present on said nucleotide has been incorporated into extended primers;
- 20
- e) repeating steps c) and d) one or more times.

25 This embodiment of the present invention can reduce costs, since few labelled nucleotides are needed. It can also be used to reduce quenching effects, should this be a problem if certain labels are used where the signals from such labels interfere with each other.

The aspects of the other embodiments of the present invention discussed herein apply *mutatis mutandis* to the foregoing embodiment of the present invention.

5 In other embodiments of the present invention all or most of the nucleotides used can be labelled and primer extension can therefore result in the contiguous incorporation of labelled bases.

10 In one of these other embodiments, only one type of label is present on four types of nucleotides to be used in extending a primer. Each nucleotide incorporation can therefore provide a cumulative increase of the same signal (e.g. of a signal measured at a particular wavelength).

15 After several steps of sequencing during which a plurality of labelled nucleotides have been incorporated into a nucleic acid strand by primer extension, it may be desired to reduce a signal to an initial level or at least to some extent. This can be achieved by removing one or more labels from already incorporated nucleotides (e.g. by laser bleaching of the fluorophores).

20 Alternatively, polymerisation steps may proceed with another type of label from that used initially (e.g. switching from fluorescein to rhodamine).

In less preferred embodiments, different labels may be used for each type of nucleotide. For example, a different fluorophore for each dNTP may be used.

25 In other less preferred embodiments, the primer itself and its extension product may be removed and replaced with another primer. If required, several steps of

sequential label-free nucleotide additions may be performed before actual sequencing in the presence of labelled nucleotides is resumed.

5 A washing step is preferably incorporated after each primer extension step in order to remove unincorporated nucleotides which may interfere with subsequent steps. The preferred washing solution should be compatible with polymerase activity and have a salt concentration which is high enough not to interfere with the annealing of primer molecules and templates.

10 In less preferred embodiments, the washing solution may interfere with polymerase activity. Here the washing solution would need to be removed before further polymerisation occurred.

15 Various detection systems can be used to detect labels (although in certain embodiments detection may be possible simply by eye, so that no detection system is needed). A preferred detection system for fluorescent labels is a Charge-Coupled-Device (CCD) camera, possibly coupled to a magnifying device. Any other device allowing detection and, preferably, also quantification of fluorescence on a surface may be used. Devices such as fluorescent imagers
20 or confocal microscopes may be chosen.

In less preferred embodiments, the labels may be radioactive and a radioactivity detection device would then be required. Ideally such devices would be real-time radioactivity imaging systems. Less preferred are other devices relying on
25 phosphor screens (Molecular Dynamics) or autoradiography films for detection.

In embodiments of the method where template molecules are immobilised in containers (e.g. wells), sequencing reactions can be monitored for each container. This can be done using microplate readers to achieve high orders of parallelism. Such readers are available commercially for the measuring of
5 fluorescence or radioactivity (e.g. Discovery™, FluoroCount™ or TopCount™ microplate readers from Packard Instrument Company).

Depending on the number of locations where detection of signals is desired, a scanning system may be preferred for data collection. (Although an alternative
10 is to provide a plurality of detectors to enable all locations to be covered.) Such a system allows a detector to move relative to a plurality of locations to be analysed. This is useful when all the locations providing signals are not within the field of view of a detector. The detector may be maintained in a fixed
15 position and locations to be analysed may be moved into the field of view of the detector (e.g. by means of a movable platform). Alternatively the locations may be maintained in fixed position and the detection device may be moved to bring them into its field of view.

The detection system is preferably used in combination with an analysis system
20 in order to determine the number (and preferably also the nature) of bases incorporated by primer extension at each location after each step. This analysis may be performed immediately after each step or later on, using recorded data. The sequence of template molecules immobilised at a given location can then be deduced from the number and type of nucleotides added after each step.

25 Preferably the detection system is part of an apparatus comprising other components. The present invention includes an apparatus comprising a

plurality of labelled nucleotides, a nucleic acid polymerase and detection means for detecting labelled nucleotides when incorporated into a nucleic acid molecule by primer extension, the detection means being adapted to distinguish between signals provided by labelled nucleotides incorporated at different
5 locations.

The apparatus may also include temperature control, solvent delivery and washing means. It may be automated.

10 An apparatus as described above may be used in a highly preferred embodiment of the present invention. Here different template molecules can be sequenced as follows:

1. each type of DNA molecule is arrayed and covalently bound at a different location (e.g. a spot) of a flat solid surface;
- 15 2. the molecules are denatured to produce single-stranded DNA;
3. oligonucleotide primers are specifically annealed to the single-stranded DNA template molecules; unbound primers are removed;
4. a reagent mixture for nucleotide extension (containing e.g. a single type of deoxyribonucleotide triphosphate, at least part of which are
20 fluorescently-labelled, a DNA polymerase and a suitable reaction buffer) is applied to the DNA array;
5. the DNA array is washed to remove unincorporated nucleotides;
6. the amount of fluorescence at each location is measured and recorded;
- the reaction proceeds to step 4) using another type of nucleotide until a
25 sufficient amount of sequence data has been generated.

Methods and apparatuses within the scope of the present invention can be used in the sequencing of:

- unidentified template molecules (i.e. *de novo* sequencing);
- and templates which are to be sequenced to check if one or more differences relative to a known sequence are present (e.g. identification of polymorphisms). This is sometimes referred to as "re-sequencing".

For *de novo* sequencing applications, the order of nucleotides applied to a given location can be chosen as desired. For example one may choose the sequential addition of nucleotides dATP, dTTP, dGTP, dCTP; dATP, dTTP, dGTP, dCTP; and so on. (Generally a single order of four nucleotides would be repeated, although this is not essential.)

The number of steps required for *de novo* sequencing a sequence of n bases is dependent on the template sequence: a high number of identical bases repeated in a contiguous stretch reduces the number of steps, whereas a template sequence having fewer such tandem repeats increases the number of steps. Without any identical base repeated contiguously, the maximum number of *de novo* sequencing steps required is $3n+1$. (After a given base has been incorporated, the next based to be added is one of the 3 other bases. The exception to this is the first step of primer extension where the first base is any of 4 bases.) Thus, for templates without any repeated bases, the probable average number of steps required to sequence n bases is very close to $2n$ (considering that any of the 3 bases which may get incorporated at a given cycle has an equal probability to be added, thus meaning that in average a base will be added every 2 cycles). Thus, for biological templates which generally contain numerous bases in tandem, the actual number of cycles required for *de novo*

sequencing will be lower than 2 times the number of bases to be sequenced. For a truly random sequence, it can be shown that the number of steps required to sequence n bases is, in average, $1.5n$ steps.

- 5 For re-sequencing applications, the order of nucleotides to be added at each step is preferably chosen according to a known sequence.

Re-sequencing may be of particular interest for the analysis of a large number of similar template molecules in order to detect and identify sequence differences (e.g. for the analysis of recombinant plasmids in candidate clones after site directed mutagenesis or more importantly, for polymorphism screening in a population). Differences from a given sequence can be detected by the lack of incorporation of one or more nucleotides present in the given sequence at particular stages of primer extension. In contrast to most commonly used techniques, the present method allows for detection of any type of mutation such as point mutations, insertions or deletions. Furthermore, not only known existing mutations, but also previously unidentified mutations can be characterised by the provision of sequence information.

- 20 In some embodiments of the method, long template molecules may have to be re-sequenced by several sequencing reactions, each one allowing for determination of part of the complete sequence. These reactions may be carried out at different locations (e.g. each location with the same template but with a different primer), or in successive cycles (e.g. between each cycles the primers and extension products are washed off and replaced by a different primer).
- 25

The present invention will now be described by way of example only, with reference to the accompanying drawings. Unless the context indicates otherwise, techniques described with regard to DNA molecules should also be considered to be applicable to RNA molecules.

5

FIGURE 1 illustrates in schematic form the *in situ* sequencing of two different DNA molecules, each molecule being present in multiple copies at particular locations, as indicated by squares or circles.

10

FIGURE 2 illustrates how fluorescence readings can be used to determine a DNA sequence *in situ*.

FIGURE 3, 4, 5 and 6 illustrate apparatuses of the present invention.

15

FIGURES 7, 8 and 9 illustrate that *in situ* sequencing done via the present invention can be confirmed (if desired) by standard techniques using gels.

FIGURES 10 and 11 illustrate *in situ* sequencing using fluorescent labels.

Figure 1

Referring now to Figure 1, individual spots or arrayed DNA templates are provided at different locations, each comprising large numbers of identical nucleic acid molecules.

5

Figure 1(a) shows the DNA templates before fluorescence labelled nucleotides are added. In this example, two different DNA templates are represented: template 1 by squares and template 2 by circles.

10

Figure 1(b) shows the DNA templates after a fluorescently-labelled dGTP has been added in the presence of a DNA polymerase and the DNA templates have been washed to remove any labelled dGTP not used in primer extension. The DNA templates which have incorporated one or more labelled Gs are represented by the black-filled squares.

15

Figures 1(c), (d) and (e) show how the procedure illustrated by Figure 1(b) has been repeated using labelled nucleotides dATP, dTTP and dCTP respectively (and washing after each step to remove nucleotides not used in primer extension).

20

Fluorescence detection can be used to distinguish incorporation of fluorescence at each location. The latest base to be incorporated at a given location by a primer extension is illustrated in Figures 1(b) to 1(e) by different appearances for different bases (G is represented by black colour, A by oblique shading, T by white colour and C by dots).

25

Figure 1(f) shows two partial sequences which have been determined by the method illustrated in Figures 1(a) to (e). The sequence shown in lane 1 is GAC, whereas that shown in lane 2 is ATC. For ease of reference, the DNA templates shown in Figure 1(a) have been identified with numbering corresponding to the lane numbering shown in Figure 1(f). It can thus be seen that two different types of DNA template are present in Figure 1(a) (i.e. a given DNA molecule is present in DNA templates identified with "1" in substantially homogenous form and with a different DNA molecule is present in DNA templates identified with "2", also in substantially homogenous form).

Of course the present invention can be used to sequence many more than two different nucleic acid molecule sequences and can be used for sequencing RNA as well as DNA.

Figure 2

Turning now to Figure 2, *in situ* sequencing is illustrated using a particular primer annealed to a template.

Figure 2(a) shows the primer annealed to the template. The primer can be seen to have a 3' end available for primer extension.

Figure 2(b) shows how fluorescence-labelled bases complementary to the underlined bases shown in Figure 2(a) can be incorporated by a stepwise primer extension cycle using fluorescently-labelled nucleotides.

Figure 2(c) shows how bases with fluorescent labels, which can be incorporated by primer extension, can be detected *in situ* by fluorescence measurements. The graphic illustrates fluorescence intensity values at a given location.

5 **Figure 3**

Turning now to Figure 3, an apparatus is illustrated for performing the present invention.

The features shown are listed below:

10

1. Arrayed DNA samples and transparent cover plate.
2. Seal around cover plate.
3. Bottom plate.
4. Inlet for reagents. (A plurality of conduits could be provided.)
- 15 5. Electronically controlled valves.
6. Reservoir of reagents.
7. Detection device (e.g. CCD camera).
8. Outlet for reagents.
9. Electronically controlled valves.
- 20 10. Reservoir of reagents. N.B. The reagents can be recycled directly into reservoir 6.

In use, template molecules are arrayed onto a plate and are covered by a transparent cover (*e.g.* a quartz plate and a cover slip may be provided). The sides of the plate and cover are sealed so that liquid can be circulated between
25 them. The reagents, which can be recycled, are distributed via a system of conduits and valves controlled electronically. Ideally a computer is used to

program the sequential incubation of the samples with appropriate reagents under appropriate conditions. The same or another computer may also control the detection system (*e.g.* a CCD camera and possibly means for moving the camera relative to the template molecules) and perform data analysis.

5

Figure 4

Figure 4 illustrates another apparatus for performing the present invention. This apparatus is preferred for manually controlled stepwise sequencing. The features illustrated are listed below:

10

1. Arrayed DNA samples.
2. Deformable absorbent material containing reagents, including appropriate type of nucleotide, to be used in primer extension.
3. Washing device.

15

In use, the deformable absorbent material (*e.g.* agarose gel or cellulose fibres) holds sequencing reagents (*e.g.* polymerase, buffer, and one type of nucleotide at a time). A soft pressure of the deformable absorbent material onto the surface where the template molecules are arrayed can liberate enough reagents for a primer extension reaction to occur, with minimal reagent waste. A washing step may then be performed by causing liquid to flow over the arrayed DNA samples, using the washing device.

20

Figure 5

25

An automated system, as shown in Figure 5, can also be designed which uses the deformable absorbent material and reagents held thereon. Ideally, four types of reagent mixes (one for each type of nucleotide) are provided by separate

areas of the deformable absorbent material. The deformable absorbent material may be generally cylindrical in shape and may have four different regions provided on the outer surface of a cylinder. The template molecules may be arrayed on the surface of another cylinder. The template cylinder may be of smaller diameter (e.g. $dt = 1/4dr$ where dt is the diameter of the template cylinder and dr is the diameter of the reagents cylinder) or larger diameter (e.g. $dt = dr + 1/4$ or $dt = dr + 3/4$). A continuous cycling of reactions can be achieved by rolling the reagent and sample cylinders against each other. Washing and detection can also be performed continuously along the surface of the template cylinder.

The automated system illustrated by Figure 5 will now be described in greater detail:

(Figure 5A shows a perspective view of the system. Figure 5B shows a plan view of the system.)

The components are as follows:

- 1.: Cylinder covered with deformable absorbent material containing sequencing reagents. Shown here is an example of cylinder with 4 sectors, which may each contain a different reagent (e.g. polymerisation mixtures, each one with a different nucleotide).
- 2.: Cylinder of smaller diameter (e.g. one fourth of the diameter of cylinder 1, as shown here) on which the samples are arrayed.
- 3.: Washing device.
- 4.: Detection device (e.g. CCD camera).

Typically, both cylinders roll against each other, so that the reagents on cylinder 1 are applied on surface of cylinder 2. The order of reagents application would correspond to their disposition on cylinder 1.

- 5 It is possible to modify this order of reagent application in other embodiments of the invention (e.g. for re-sequencing). For example, an additional device can be used to skip application of one or more reagents onto cylinder 2. This can be achieved by controlled rotation of cylinder 1 relative to cylinder 2 while contact between both cylinders is avoided.

10

Figure 6

- A further apparatus, as shown in Figure 6, can be designed. Reagents can be dispensed by microspraying. Thus minimal volumes of reagents need be used at each step. For instance, automated piezoelectric devices can be designed to
- 15 dispense nanolitres (or picolitres) of reagents on defined locations (e.g. ink-jet printing technology). The re-sequencing of different templates, or of similar templates from different primers, can be achieved by dispensing the appropriate reagents (e.g. primer in annealing solution or polymerase, buffer and nucleotide) at defined locations.

20

Three systems that can be used successively to carry out a sequencing reaction are illustrated in Figure 6:

- A. Piezoelectric device for dispensing the appropriate reagents at defined locations. The device is linked to an electronic controller and to a reservoir of
- 25 reagents.
- B. Washing device.
- C. Detection device.

Figure 7

Illustration of the sequencing method with ^{32}P -labelled primer, as described in Example 1. The template was immobilised on magnetic beads. The sequencing cycles were carried out using the indicated nucleotides in an arbitrary order (A, T, G, C,..) to mimic *de novo* sequencing. The reaction products were visualised by autoradiography after electrophoresis on denaturing acrylamide gel. Fragment sizes are indicated on the right with the predicted sequence.

Figure 8

Illustration of the sequencing method with ^{32}P -labelled nucleotides, as described in Example 2. The template was immobilised on magnetic beads. The sequencing steps were carried out using the indicated nucleotides in the order of expected incorporation. The primer was unlabelled. The reaction products were visualised by auto-radiography after electrophoresis on denaturing acrylamide gel. Fragment sizes are indicated on the right with predicted sequence.

Figure 9

Illustration of the sequencing method without gel electrophoresis, as described in Example 3. The sequencing steps were carried out using the indicated ^{32}P -labelled nucleotides in the order of expected incorporation. The primer was unlabelled. Radioactivity accumulation in each well was recorded after each cycle and is shown on the graphic (left scale) with the predicted number of added bases (right scale).

Figure 10

Illustration of the sequencing method with fluorescence-labelled nucleotides, as described in Example 4. The template was immobilised in a microplate and the primer was unlabelled. The nucleotides were used in the order of T, C, T, C, T, C, T, which corresponds to the expected sequence. The fluorescence increase after each step was recorded (arbitrary units) and is shown on the graph with the expected number of added bases. The first step corresponds to non specific signal.

Figure 11

Illustration of parallel sequencing with fluorescently-labelled nucleotides, as described in Example 5. The template was immobilised on a microplate and different unlabelled primers were used on two locations. The nucleotides were used in the arbitrary order of T, C, T, C on both locations. The fluorescence increase after each step was recorded for each location and is shown on the graphic. The predicted number of bases added to each primer is shown in Table 11.

EXAMPLES

Example 1

5 This example provides an indication that successful step-by-step incorporation of nucleotides using the method of the present invention can be achieved. The example uses a sequencing gel for illustrative purposes (see Figure 7) since such gels are routinely used for sequencing. It will however be appreciated that the present invention is preferably used for *in situ* sequencing of nucleic acid
10 molecules and therefore it does not require the use of sequencing gels.

Experimental procedure:

1. The sequencing cycles were done on single-stranded template molecules
15 bound on magnetic beads.

The 174 base pair DNA fragment was amplified by PCR from the multiple cloning site of plasmid pBlueScript SK- (Stratagene). The forward primer (5'-biotin-GCGCGTAATACGACTCACTA-3') and
20 reverse primer (5'-CGCAATTAACCCTCACTAAA-3') were located on position 621 and 794, respectively. The amplified double-stranded DNA was bound on streptavidin coated magnetic beads (Dynabeads M-280, Dynal, Oslo) and denatured with NaOH. After washing, the single-stranded DNA bound to the beads was kept in 10 mM tris pH 8.0.

25

2. The reverse primer was used for sequencing. It was labelled on its 5' end with radioactive phosphate, and annealed to the single-stranded DNA template.

The primer molecule is expected to be extended with the following bases :
GGGAACAAAAGCTGGAG...

5 The primer end-labelling was done with polynucleotide kinase
(Pharmacia) in presence of [γ -32P] ATP as described by manufacturer.
After purification on Sephadex G-25 spin column, the primer were
conserved at -20°C.

10 The annealing was performed by heating the mixture of primer and
template molecules in 1x Sequenase buffer [40 mM Tris pH 7.5, 20 mM
MgCl₂, 50 mM NaCl] for 5 min at 70°C and slow cooling for 2 hours.

3. The sequencing reaction were performed by successive cycles of primer
extension (in presence of only one type of nucleoside triphosphate), aliquot
15 removal for analysis and washing. Each cycle was done using a different
deoxyribonucleoside triphosphate in solution, using the arbitrary order of
dATP, dTTP, dGTP, dCTP, dATP, dTTP, dGTP, dCTP, dATP, dTTP, etc.

20 The reaction mix contained 0.1 u/ μ l Sequenase (T7 Sequenase II,
Amersham) and 100 nM of a single type of normal deoxyribonucleoside
triphosphate in 1x Sequenase buffer. Four different mixes were
prepared, each with a different type of nucleoside triphosphate.

25 A cycle started with the removal of buffer from the template-bound
beads using a magnet. Ice-cold reaction mixture was added. After 5
seconds an aliquot was collected and kept on ice in 10 volumes of stop
buffer containing 90% formamide and 20 mM EDTA. Ten volumes of

washing buffer (10 mM Tris pH 8.0) were added to the remaining reaction mixture and immediately removed after magnetic separation of the template-bound beads. Washing was repeated and next cycle was initiated.

5

At each cycle, a smaller volume of fresh reaction mixture was used, according to the remaining amount of template after aliquot removal and washing losses.

- 10 4. The aliquots collected after each cycle were then analysed on denaturing sequencing gel. After migration, the gel was autoradiographed to reveal to position of each band.

15 The aliquots in stop buffer were denatured for 5min at 95°C and immediately cooled on ice before loading on 0.75 x 180 x 320 mm 15% acrylamide gels (SE420, Hoefer Scientific Instruments) containing 1xTBE buffer and 7 M urea. Migration was at 2000V for 5min and 250V overnight.

20 As a control, unreacted oligonucleotides were applied on gel (lane 1), as well as the product of a reaction performed in presence of all 4 deoxynucleoside triphosphates, to give rise to fully extended molecules (lane 2).

25 After migration, the gel was fixed in 10% glacial acetic acid and 10% methanol for 30 min. Still humid, it was transferred in a plastic bag in

autoradiography cassette to expose a Kodak X-OMAT AR film for 9h at 25°C.

5 The very high sensitivity of radio-active label, associated with gel analysis of extended fragments after each cycle allowed careful monitoring of the reaction. As shown on Figure 7 it was demonstrated that:

- 10 • Conditions were identified which permit correct polymerisation in the presence of only one type of deoxynucleoside triphosphate. The polymerisation proceeds completely to the last base to be incorporated (lane 5), but is blocked for (further) extension if the complementary base is not present on template molecule (lane 3, 4, or 5).
- 15 • There is no misincorporation, as seen on lanes 3, 4, or 6.
- There is no 3'-5' exonuclease activity (e.g. fragments of shorter size are not detected on lane 3 or 4).
- 20 • Successive cycles can be carried out correctly.

25

Only fragments of the correct size are observed. There are three exceptions : (1) a small proportion of the primer molecules were not extended during first cycle, probably because of 3' end damage of the oligonucleotide. (2) A proportion of full length molecules are observed after cycle 13, which suggests that washing was not complete and that the 4 deoxynucleoside triphosphates were present in solution. This

problem is not considered as relevant, because it was not observed between cycles 4 and 12, even though the 4 different deoxynucleosides triphosphates had already been used for reactions. The washing steps should also be facilitated when the experiment will be repeated with
5 template molecules bound on plastic surface. (3) Faint bands of smaller size than expected could be detected after longer exposure, which suggests uncompleted reaction. However, the proportion is so low that it should not interfere with *in situ* detection.

10 Only 19 cycles were carried out in this preliminary experiment, but there is no reason why more cycles (up to 50 or up to 100 or more) could not be performed successfully.

This experiment demonstrated that not only re-sequencing (when only the
15 expected deoxynucleoside triphosphate is used for polymerisation), but that also *de novo* sequencing can be carried out efficiently with this method.

After 20 cycles of *de novo* sequencing of an unknown template molecule, the
20 minimal number of bases which could be sequenced is of 6, the maximal number depending on the base order and on the presence of runs of the same base. In this *de novo* sequencing experiment, as many as 17 bases were read in 19 cycles.

It should also be mentioned that the template molecule used here is considered
25 as difficult to sequence because of its high content of palindromic regions, which gives rise to secondary structures in single-stranded DNA. This potential difficulty appeared not to be a problem in our experiment.

Example 2: Re-sequencing using non-labelled primers

5 This example provides a further indication that successful step-by-step incorporation of labelled deoxynucleoside triphosphates can be achieved using the method of the invention. Like in Example 1, a sequencing gel was used for illustrative purposes (Figure 8). It will however be appreciated that the present invention is preferably used for *in situ* sequencing of nucleic acid molecules and therefore it does not require the use of sequencing gels.

10

Template preparation:

A 608 base pair DNA fragment inserted into the polylinker of pBlueScript SK-plasmid was amplified by PCR. The forward primer (5'-GCG CGT AAT ACG ACT CAC TA-3') was biotinylated on its 5'-end and the reverse primer (5'-GCA
15 ATT AAC CCT CAC TAA A-3') was not functionalised (i.e. -OH). The amplified double-stranded DNA was purified on a spin column (Pharmacia MicroSpin S-400 HR) to remove unused primers. A hundred microlitres of purified DNA were then diluted in a total volume of 400 µl and bound on same volume of streptavidin coated magnetic beads (10 mg/ml) according to the
20 manufacturer's protocol (Dynabeads M-280, Dynal, Oslo). The DNA was denatured in presence of 0.1 N NaOH and beads were washed to remove the complementary strand. The single-stranded DNA bound to the beads was resuspended in 100 µl 10 mM tris pH 8.0 and conserved at 4°C.

25

Primer annealing:

The primer 5'-TAC CAG TTT CCA TTC CAG C-3' was used for sequencing. The annealing was performed by heating the primer and template mixture for 5

min at 95°C and cooling to 60°C in 30 min. Typically, 20µl of beads were annealed with 10 pmol of primer in 100µl of 1x Sequenase buffer (40 mM Tris pH 7.5, 20 mM MgCl₂, 50 mM NaCl). The primer was expected to be extended with the following bases: CGCTGGGGTGGTTT... which are complementary to the template sequence.

The primer extension reactions were performed in presence of only one type of deoxynucleoside triphosphate. Each step was done using a different deoxyribonucleoside triphosphate in solution, using the expected order of dCTP, dATP, dTTP, dGTP, dCTP, dTTP, dGTP, dTTP, dGTP, dTTP, etc..(the first dATP and dTTP will not get incorporated and serve to demonstrate absence of misincorporation). The T7 DNA polymerase Sequenase 2.0 (Amersham) was chosen for polymerisation reactions. The reaction mix contained 0.2 u/µl Sequenase, 5mM DTT, 250 nM of a single type of normal deoxyribonucleotide and 5 nM of the same type of [α -³²P]-labelled deoxyribonucleotide (50 nCi/µl, Amersham) in 1x Sequenase buffer. Four different mixes were prepared, each with a different type of deoxynucleoside triphosphate.

The first polymerisation step was initiated by sedimenting 20 µl of template-bound beads. The beads were resuspended in 25 µl of ice-cold reaction mix and reaction was allowed to proceed for 10 seconds.

The beads were immediately sedimented with the magnet and washed 3 times for 1 min in 10 volumes of 1x Sequenase buffer. After the final wash, the beads were resuspended in the initial volume of 1x Sequenase buffer. One microlitre aliquot was collected after each cycles and kept in 10 µl stop buffer (90%

formamide, 1x TBE buffer, 20 mM EDTA, 0.1% Bromophenol Blue, 0.1% Xylene Cyanol FF).

5 The next polymerisation and washing steps were performed as described *supra* using adequate volumes of reagents (i.e. each step using smaller volumes to compensate for aliquot removal).

10 After final washing step, sample were analysed by gel electrophoresis. The aliquots in stop buffer were denatured for 5 min at 95°C and immediately cooled on ice before loading on 0.75 x 180 x 320 mm 15% acrylamide gel containing 1x TBE buffer and 7 M urea. The electrophoresis system was model SE400 from Hoefer Scientific Instruments. Migration was at 2000 V for 5 min and 250 V overnight. After migration, the gel was immediately transferred into a plastic bag to expose a Kodak X-OMAT AR film in an autoradiography
15 cassette for 3 to 16h at 25°C.

As discussed *supra* and shown in Figure 8, this example indicates that the sequencing method of the invention can be successfully adapted for stepwise sequencing from non-labelled primers.

20

Example 3: Sequencing template bound in a microplate well

Template preparation:

25 In this example, DNA molecules of approximately 700 bases were used as template. They were bound on the plastic surface of a modified microtitre plate as described by the manufacturer (NucleoLink™ from Nunc A/S, Roskilde, Denmark).

Two types of DNA templates (named A and B) were bound to the wells, either individually or mixed (50% each).

Primer annealing:

- 5 A 20 base-long oligonucleotide was used as sequencing primer (5'-GGT CAG GCT GGT CTC GAA CT-3') specific for DNA template B. The annealing was performed by heating the microtitre plate coated with template for 4 min at 94°C and slow cooling to 60°C in 30 min, in presence of 100nM primer (in 20µl 5xSSC buffer +0.1% Tween® 20 per well). The primer was expected to be
- 10 extended with the following bases: CCCTACCTCA... which are complementary to the template sequence.

Polymerisation step:

- The primer extension reactions were performed in presence of only one type of
- 15 deoxynucleoside triphosphate. Each step was done using a different deoxynucleoside triphosphate in solution, using the expected order of dCTP, dTTP, dATP, dCTP, dTTP, dCTP, dATP. The Klenow fragment of DNA polymerase I (New England Biolabs) was used for polymerisation reactions. The reaction mix contained 0.25 u/µl Klenow exo-, 50 nM of a single type of
- 20 normal deoxyribonucleoside triphosphate and 20 nM of the same type of [α -³²P]-labelled deoxyribonucleoside triphosphate (50 nCi/µl, Amersham) in 1x Polymerase I buffer (10 mM Tris pH 7.5, 5 mM MgCl₂, 7.5 mM dithiothreitol). Four different mixes were prepared, each with a different type of
- deoxynucleoside triphosphate. Typical reactions were performed in 12 µl.

25

A polymerisation step was initiated with the removal of reagents from the well. The adequate volume of reaction mixture was added and reaction was allowed to proceed for 30 seconds at room temperature.

5 **Washing step:**

The wells were immediately washed 3 times for 1 min in 100 μ l of TNT buffer (100 mM Tris pH 7.5, 150 mM NaCl, 0.1% Tween® 20).

Detection step:

10 After each washing step, the radioactivity incorporated on each well was measured and recorded using a scintillation counter.

Sequencing cycle:

15 In this example, the polymerisation, washing and detection steps were repeated sequentially.

Results shown in Figure 9 indicate that primer extension occurred as expected. The stepwise counts increase occurred specifically in presence of template B, and was dependent on its amount bound on wells (e.g. approximately twice as
20 much counts in presence of 100% template B than in presence of 50% template B). After each step, the number of counts detected increased almost proportionally to the number of bases added.

Example 4: Sequencing using fluorescent label

25

This example demonstrates that sequencing can be achieved *in situ* by stepwise incorporation of fluorescently-labelled nucleoside triphosphates. The number of

added bases is determined by the measurement of the fluorescence increase after each cycle of addition. Thus DNA sequencing is achieved without requirement of electrophoresis separation method.

5 **Template preparation:**

In this example, a 157 bp DNA fragment was bound to a streptavidin-coated 96-well microtitre plate. The fragment was prepared by PCR amplification with the forward primer being 5'-biotinylated. The reverse primer was not functionalised. The unincorporated primers were removed by purification of the PCR product on a spin column (Pharmacia MicroSpin S-400 HR). After binding to the streptavidin coated microtitre plate wells (LabSystems, Helsinki), the PCR fragments were denatured by NaOH treatment. After washing, only the forward strand remained bound to the wells.

15 **Primer annealing:**

The primer 5'-ACA CGA GCC ACA CTG TCC CAG GGG C-3' was used for sequencing. The annealing was carried out at 25°C with 0.5 µM of primer in 5x SSC buffer.

20 **Primer extension:**

The expected primer extension on the template is : C, T, C, C, T, C, T. The reaction mix [1x Sequenase buffer, 0.15 u/µl T7 DNA polymerase (Sequenase 2.0, Amersham), 0.1 % BSA, 5 mM DTT] contained 0.5 µM of dUTP or dCTP [as a 60% mix of Cy5-labelled (Amersham) and non-labelled deoxynucleotide triphosphates]. The reactions were carried out for 1 min at 25°C in presence of either dUTP or dCTP. After each step, the wells were washed twice with 1x Sequenase buffer and the fluorescence intensity in each well was measured with

a cooled CCD camera (Princeton Instruments) mounted on a fluorescence microscope (Zeiss Axiovert TV100). Note that dTTP can be substituted by dUTP, the two nucleotides having identical specificity.

- 5 Results shown on Figure 10 demonstrate that after each cycle of dNTP addition, the measured increase in fluorescence is proportional to the number of bases incorporated. The successive reactions were performed alternatively in presence of dUTP or dCTP. dUTP was used for the first step, which served as a control of non-specific extension.

10

In this example, 8 bases were sequenced using the method of the invention.

Example 5: Parallel sequencing using fluorescently-labelled dNTPs

- 15 This example demonstrates that *in situ* sequencing by stepwise incorporation of fluorescently-labelled deoxynucleoside-triphosphates can be achieved in parallel in different locations. Thus several samples can be sequenced simultaneously.

- 20 Template preparation:

In this example, a 131 bp DNA fragment was bound to a streptavidin-coated 96-well micro-titre plate. The fragment was prepared by PCR amplification with the forward primer being 5'-biotinylated. The reverse primer was not functionalised. The unincorporated primers were removed by purification of the

- 25 PCR product on a spin column (Pharmacia MicroSpin S-400 HR). After binding to the streptavidin coated microtitre plate wells (LabSystems, Helsinki), the

PCR fragments were denatured by NaOH treatment. After washing, only the forward strand remained bound to the wells.

Primer annealing:

- 5 Two primers were annealed to the template on different locations. Primer 79 (5'-GGG GTT TCT CCA TGT TGG TCA GGC TGG TC-3') was annealed at a first location on the microtitre plate and primer 94 (5'-TGG TCA GGC TGG TCT CGA ACT CCC TAC-3') at a second location along the microtitre plate. The annealing was carried out at 25°C with 0.5 μ M of primer in 5x SSC buffer.

10

Primer extension:

The expected extension for primer 79 is:- TCG... and for primer 94:- CTCA...

- The reaction mix [1x Sequenase buffer, 0.15 u/ μ l T7 DNA polymerase (Sequenase 2.0, Amersham), 0.1 % BSA, 5 mM DTT] contained 0.5 μ M of
15 dUTP or dCTP [as a 60% mix of Cy5-labelled (Amersham) and non-labelled deoxynucleotides]. The reactions were carried out for 1 min. at 25°C in presence of either dUTP or dCTP. After each step, the wells were washed twice with 1x Sequenase buffer and the fluorescence intensity in each well was measured with a cooled CCD camera mounted on a fluorescence microscope, as
20 described above.

- Results shown on Figure 11 demonstrate that after each step and for each primer the fluorescence increase is proportional to the number of bases incorporated. The successive reactions were performed alternatively in presence
25 of dUTP or dCTP. Fluorescence increased on steps 1 and 2 for primer 79, as one base was added on these steps (see table). For primer 94 however, bases were only incorporated on steps 2, 3, and 4. As expected, no increase of

fluorescence was detected for primer 79 on steps 3 and 4, and for primer 94 on step 1.

5 The example shows the specificity of the method and its use for sequencing in parallel simultaneously on different locations.

Use of double stranded, nicked molecules

10 Although the foregoing examples have been generally based upon a method in which primers are hybridised to a single-stranded nucleic acid molecule and the primers are then extended, it is possible to use an alternative approach in which a double stranded molecule having a nick is provided (a nick being a gap in one strand of a double-stranded DNA molecule and allowing a free 3'-OH end to be provided).

15 The nick provides a 3'-OH group which acts as a starting point for chain extension in the presence of a suitable polymerase and a supply of nucleotides. The 3'-OH group provided via a nick therefore has an equivalent function to the 3'-OH group at the end of a primer in the methods discussed previously. The
20 nick can be provided by any suitable means. For example it can be provided by using restriction enzymes on DNA molecules with hemiphosphorothioated recognition sites. (Reference : Spears, P.A., Linn, C.P., Woodard, D.L., Walker, G.T. (1997). Simultaneous strand displacement amplification and fluorescence polarization detection of Chlamidia trachomatis DNA. Analytical
25 Biochemistry 247: 130-137).

Claims

1. A method for sequencing nucleic acid molecules, comprising the steps
of:
- 5
- a) providing at a first location a plurality of single stranded nucleic acid
molecules that have the same sequences as one another and that are
hybridised to primers in a manner to allow primer extension in the
presence of nucleotides and a nucleic acid polymerase;
- 10
- b) providing at a second location, which is different from the first location,
a plurality of single stranded nucleic acid molecules that have the same
sequences as one another, but that have different sequences from the
sequences of the single stranded nucleic acid molecules at the first
15 location, and that are also hybridised to primers in a manner to allow
primer extension in the presence of nucleotides and a nucleic acid
polymerase;
- 20
- c) providing each location with a nucleic acid polymerase and a given
labelled nucleotide under conditions that allow extension of the primers
if a complementary base or if a plurality of such bases is present at the
appropriate position in the single stranded nucleic acid molecules;
- 25
- d) detecting whether or not said labelled nucleotide has been used for
primer extension at each location by determining whether or not the label
present on said nucleotide has been incorporated into extended primers;

- e) repeating steps c) and d) one or more times so that extended primers comprising a plurality of labels are provided.
2. A method according to claim 1, wherein all or part of the sequence that
5 is obtained in step e) is converted to provide a complementary sequence thereto.
3. A method according to any preceding claim, wherein if the given
nucleotide has been used in primer extension in step d) then this step includes
the step of detecting how many of the given nucleotides have been used per
10 extended primer.
4. A method according to any preceding claim, wherein after step c) excess
nucleotides that have not been used in primer extension are removed (e.g. by
washing).
15
5. A method according to any preceding claim, wherein step d) uses
absorption or emission spectrometry.
6. A method according to any preceding claim, wherein said single
20 stranded nucleic acid molecules, said primers or both of the aforesaid are
immobilised.
7. A method according to any preceding claim that is used to fully or
partially sequence 10 or more nucleic acid molecules having different
25 sequences simultaneously.

8. A method according to any preceding claim that is used to fully or partially sequence 100 or more nucleic acid molecules having different sequences simultaneously.
- 5 9. A method according to any preceding claim that is used to fully or partially sequence 1000 or more nucleic acid molecules having different sequences simultaneously.
- 10 10. A method according to any preceding claim, wherein each of four different nucleotides is used in primer extension.
11. A method according to claim 10, wherein said four different nucleotides are used in a predetermined order in repeated cycles.
- 15 12. A method according to claim 10 or claim 11, wherein the nucleotides are dATP, dTTP, dGTP and dCTP in labelled form.
13. A method according to claim 10 or claim 11, wherein the nucleotides are ATP, UTP, GTP and CTP in labelled form.
- 20 14. A method according to any preceding claim, wherein the detection step is carried out without moving the nucleic acid molecules from the different locations.
- 25 15. A method as described in any preceding claim with the exception that double stranded nucleic acid molecules having nicks therein are provided at the

first and/or second locations instead of providing single stranded molecules hybridised to primers.

16. A method as described in any preceding claim with the exception that
5 only one nucleic acid molecule is provided at the first and/or second locations.
17. A method for sequencing nucleic acid molecules, comprising the steps of:
- 10 a) providing at a first location a plurality of single stranded nucleic acid molecules that have the same sequences as one another and that are hybridised to primers in a manner to allow primer extension in the presence of nucleotides and a nucleic acid polymerase;
- 15 b) providing at a second location, which is different from the first location, a plurality of single stranded nucleic acid molecules that have the same sequences as one another, but that have different sequences from the sequences of the single stranded nucleic acid molecules at the first location, and that are also hybridised to primers in a manner to allow
20 primer extension in the presence of nucleotides and a nucleic acid polymerase;
- c) providing each location with a nucleic acid polymerase and a given nucleotide in labelled and unlabelled form under conditions that allow
25 extension of the primers if a complementary base or if a plurality of such bases is present at the appropriate position in the single stranded nucleic acid molecules;

d) detecting whether or not said labelled nucleotide has been used for primer extension at each location by determining whether or not the label present on said nucleotide has been incorporated into extended primers;

5

e) repeating steps c) and d) one or more times.

18. An apparatus for performing a method according to any preceding claim, the apparatus comprising a plurality of nucleotides, a nucleic acid polymerase and detection means for performing step d) of claim 1 or an equivalent step for
10 any of claims 15 to 17, the detection means being adapted to distinguish between said different locations.

19. An apparatus according to claim 18 comprising means for removing
15 excess nucleotides from the first and second locations (e.g. washing means).

20. An apparatus according to claim 19 or claim 18 that is automated to allow repeated cycles of primer extension and detection.

20 21. A method of sequencing a target nucleic acid comprising:
(a) hybridizing the target nucleic acid to a primer whereby the target nucleic acid can serve as a template for extension of the 3' end of the primer;
(b) incubating the hybridized target nucleic acid/primer with a
polymerase and a type of nucleotide bearing a label under conditions supporting
25 template-directed extension of the primer if the nucleotide type can be incorporated as the complement of a corresponding nucleotide of the target;

(c) measuring first label incorporated into the primer to determine whether, and if so, by how many base increments, the primer has been extended by incorporation of the nucleotide type;

5 (d) incubating the hybridized primer/target nucleic acid with a different type of nucleotide bearing a label under conditions supporting template-directed extension of the primer if the different nucleotide type can be incorporated so as to be complementary to a corresponding nucleotide in the target;

10 (e) measuring incremental label incorporated into the primer due to the previous incubating step to determine whether, and if so, by how many base increments, the primer has been extended by incorporation of the different nucleotide type; and

(f) repeating steps (b) – (e) until a desired portion of the target sequence can be determined from the incremental base additions to the primer.

15

22. A method according to claim 21 which is a method according to any of claims 1 to 14, 16 or 17.

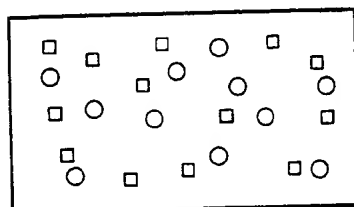
20 23. A method according to claim 21 or claim 22 with the exception that instead of hybridizing a target nucleic acid molecule to a primer and extending the primer with labelled nucleotides, a nick is introduced into a double-stranded nucleic acid molecule and the nick is extended using nick translation and labelled nucleotides.

25 24. The invention substantially as hereinbefore described.

1 / 9

FIG. 1.

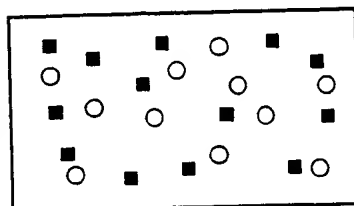
in situ Sequencing



Two types of arrayed
DNA templates
(1: squares, 2:
circles)

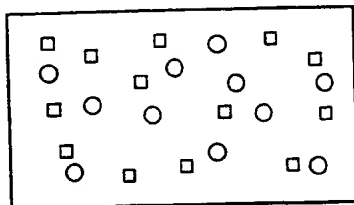
1 2

Step 1 : G



G

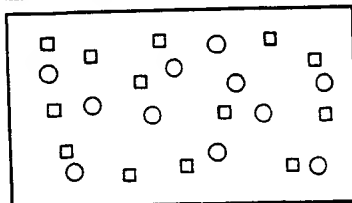
Step 2 : A



A

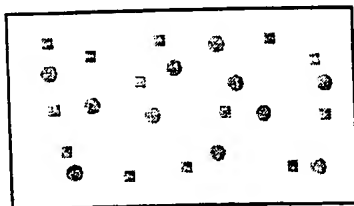
A

Step 3 : T



T

Step 4 : C



C

C

```

Sequence from      type 1 : G A C
                   type 2 : A T C

```

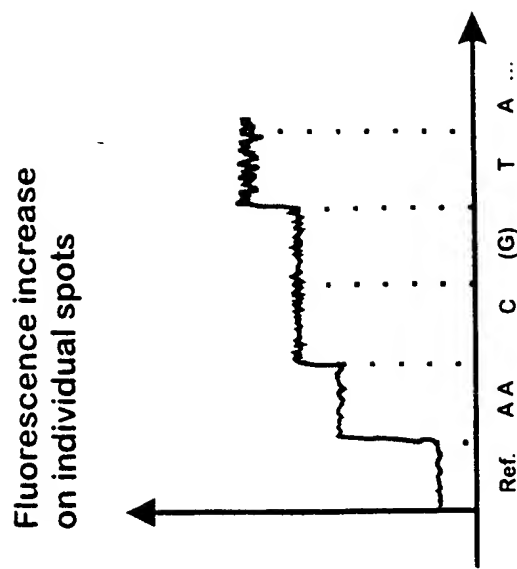
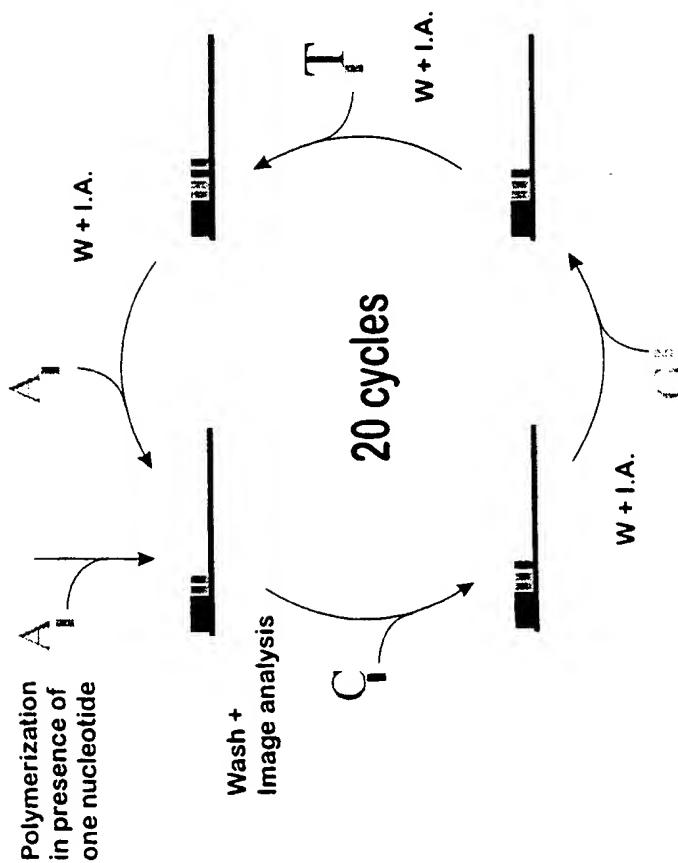
2 / 9

in situ Sequencing

FIG. 2.

Primer : 5'-gactagcgtcat-3'
 Template : 3'-ggatgctgacgcactattgatgggcacgaactca-5'

Cycle of stepwise base extension



3 / 9



FIG. 3.

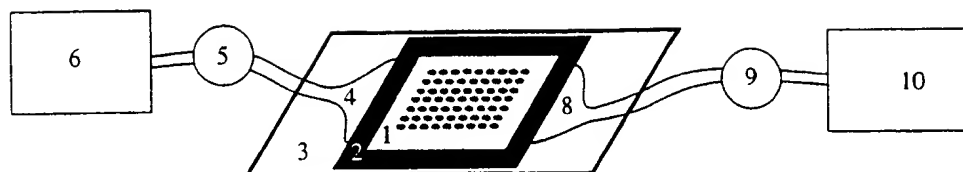
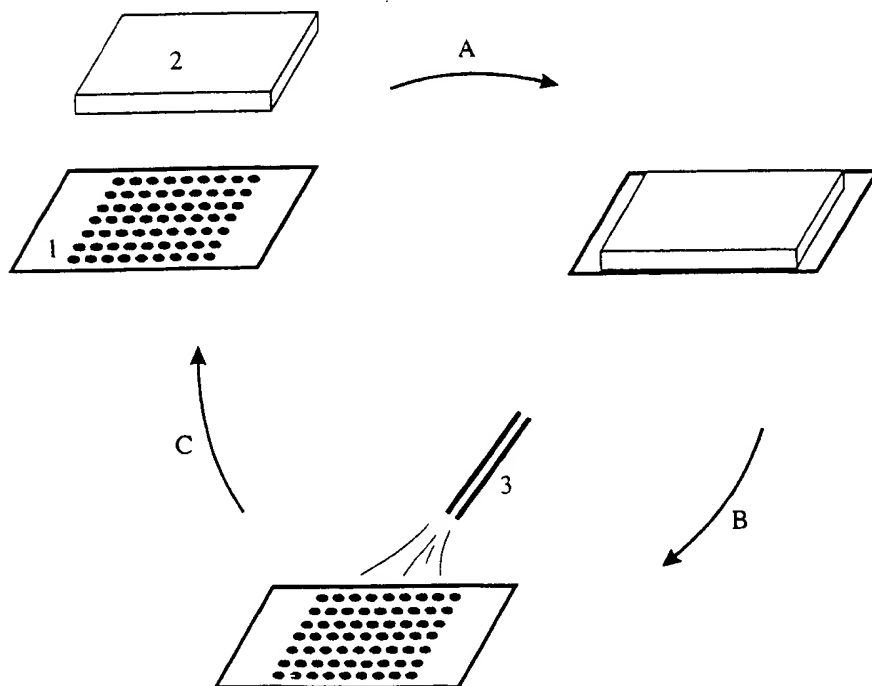


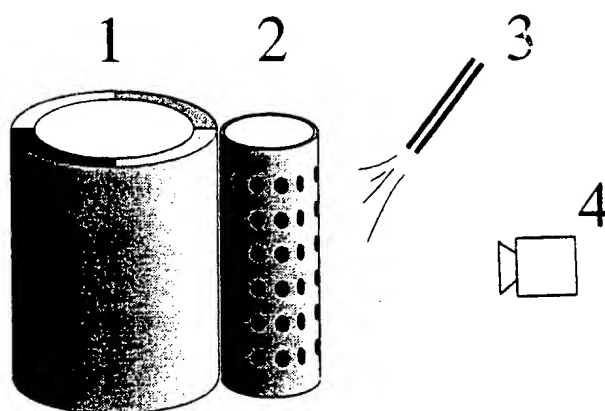
FIG. 4.



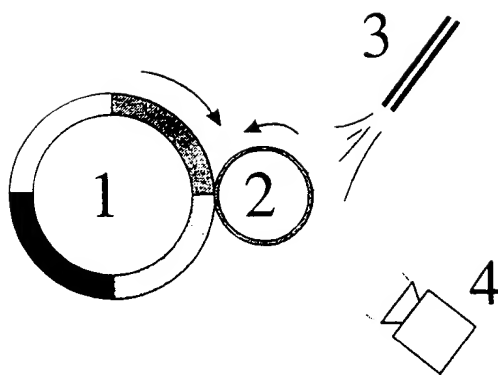
4 / 9

FIG. 5.

A.

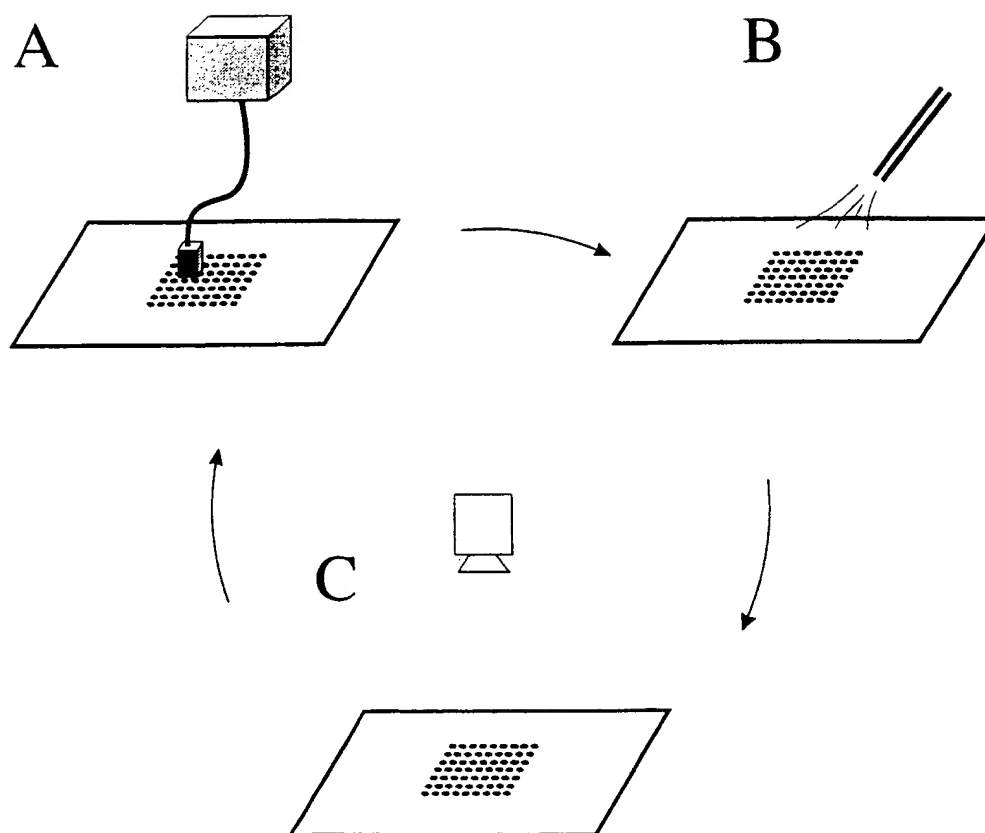


B.



5 / 9

FIG. 6.



6 / 9

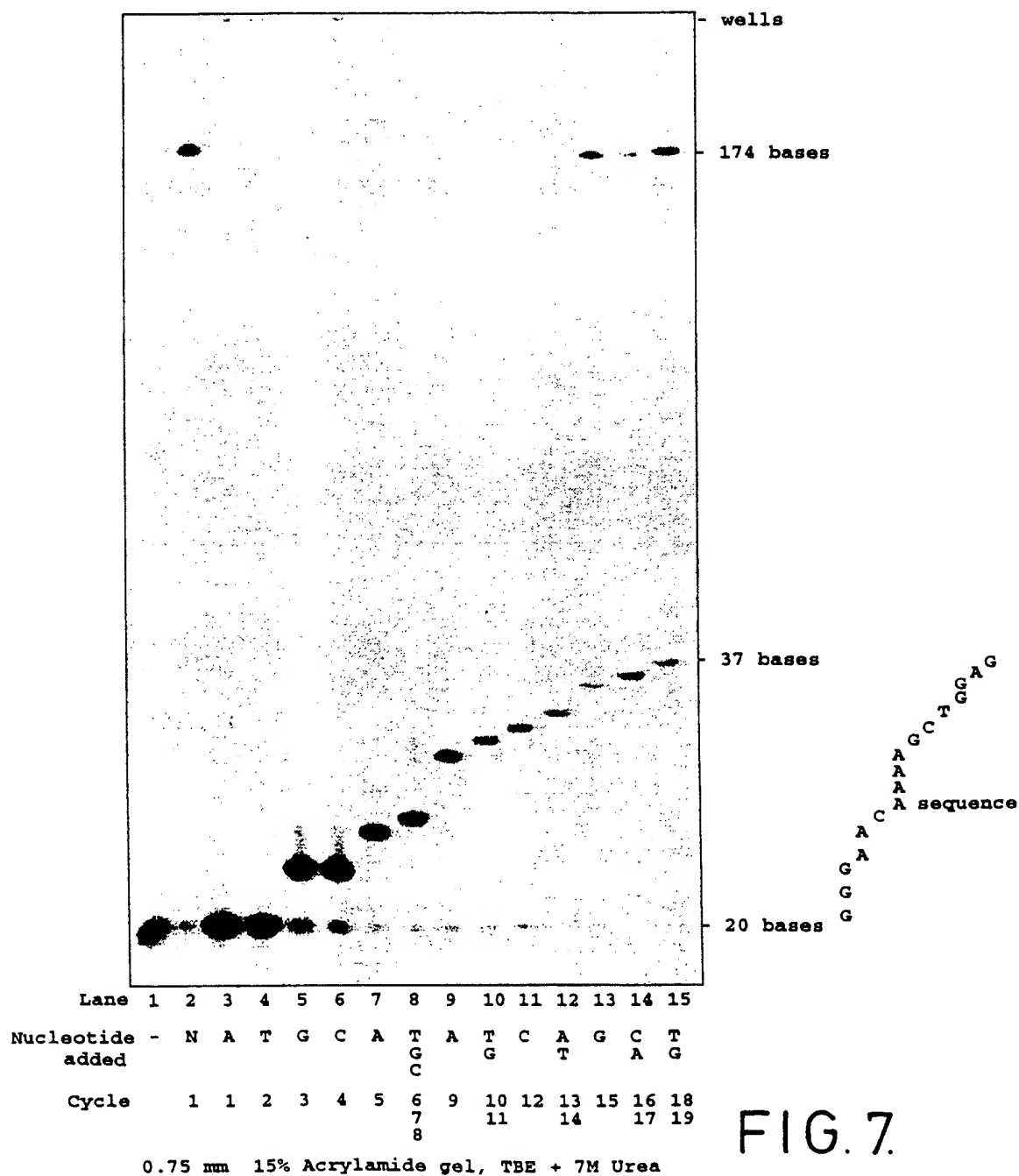
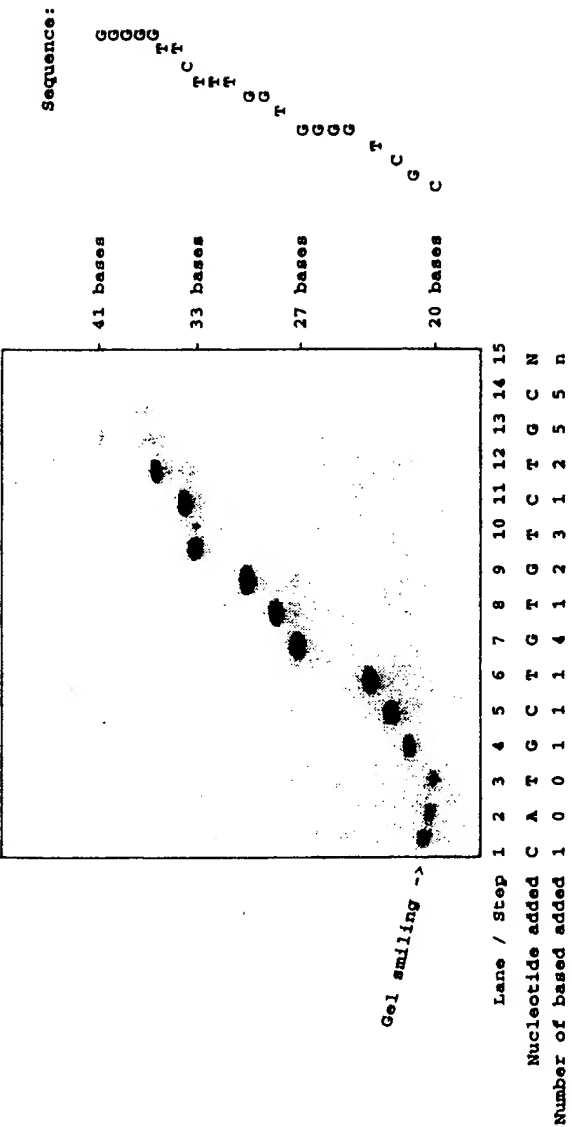
in situ Sequencing : *de novo* Sequencing

FIG.8.

in situ Sequencing : Non-labelled primers



8 / 9

FIG. 9.

Step-by-Step Primer Extension on DNA Template Bound on Microplate Wells

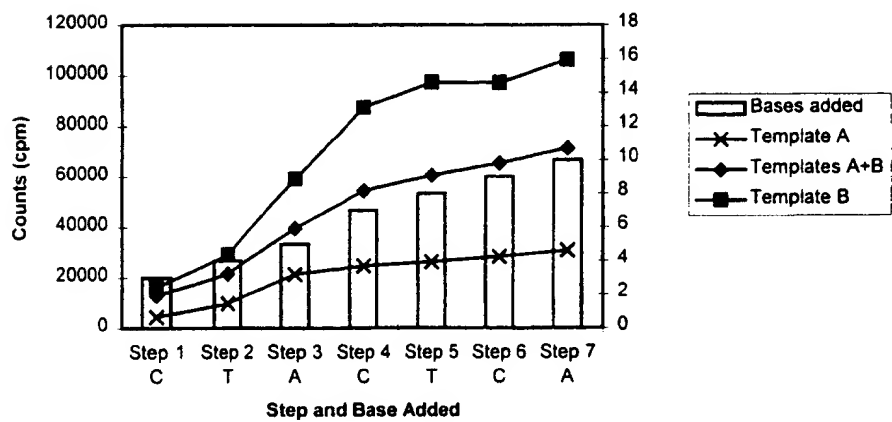
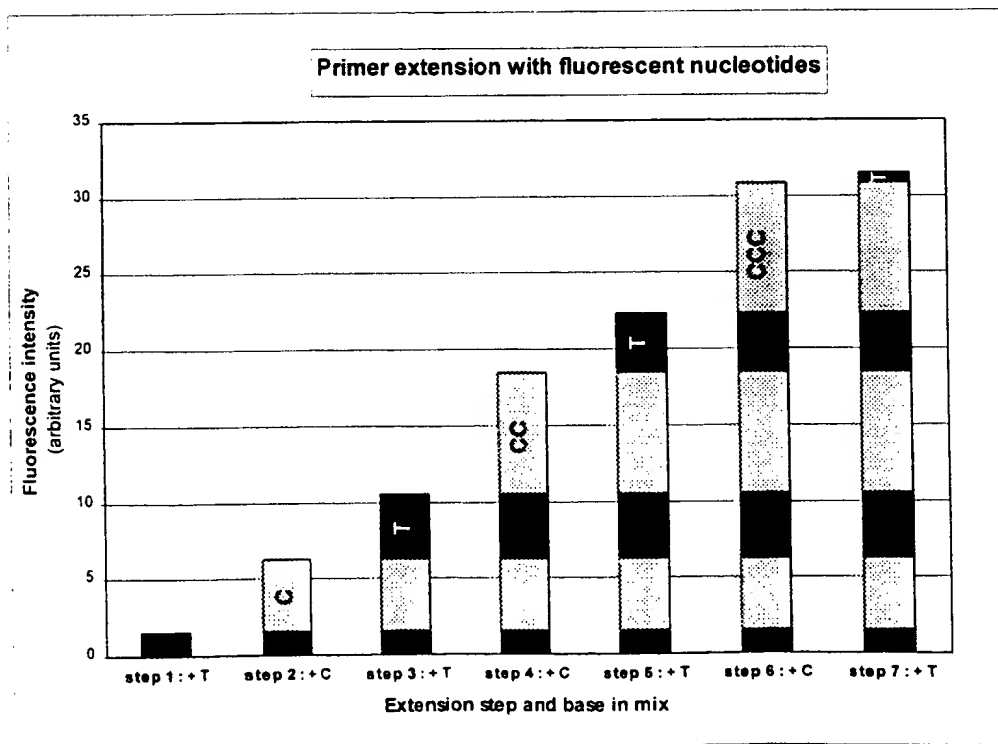


FIG. 10.



9 / 9

FIG. 11.

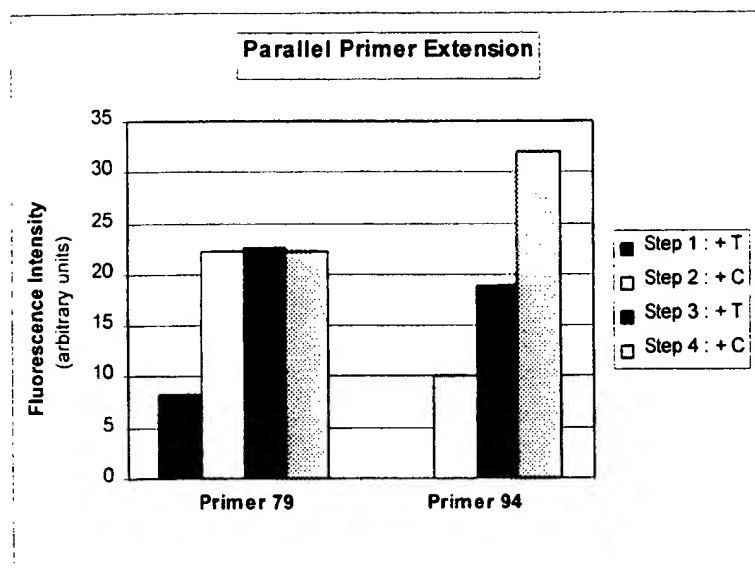


Table 11:

Step	Base in mix	Primer 79 : TCGA			Primer 94 : CTCA		
		Expected extension	Bases added	Fluorescence intensity	Expected extension	Bases added	Fluorescence intensity
1	T	T	1	8	none	0	0
2	C	C	2	22	C	1	10
3	T	none	2	23	T	2	19
4	C	none	2	22	C	3	32

INTERNATIONAL SEARCH REPORT

In International Application No

PCT/GB 98/00963

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C1201/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C120

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 665 293 A (HAMAMATSU PHOTONICS KK) 2 August 1995 see the whole document ---	1-14, 16-20
X	WO 96 27025 A (RABANI ELY MICHAEL) 6 September 1996 see the whole document ---	1-14, 16-20
X	WO 91 06678 A (STANFORD RES INST INT ;TSIEN ROGER Y (US)) 16 May 1991 see the whole document ---	1-14, 16-20
X	WO 96 32504 A (UNIV BOSTON) 17 October 1996 see the whole document ---	15,21-24
-/--		

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search

29 July 1998

Date of mailing of the international search report

05/08/1998

Name and mailing address of the ISA

European Patent Office, P. B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Hagenmaier, S

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 98/00963

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	FU D -J ET AL: "SEQUENCING DOUBLE-STRANDED DNA BY STRAND DISPLACEMENT" NUCLEIC ACIDS RESEARCH, vol. 25, no. 3, 1997, pages 677-679, XP000196977 see the whole document ---	15,21-24
A	WO 96 36737 A (RABANI ELY MICHAEL) 21 November 1996 see the whole document ---	1-24
A	DE 195 15 552 A (EUROP LAB MOLEKULARBIOLOG) 31 October 1996 see the whole document -----	1-24

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/GB 98/00963

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
EP 0665293	A	02-08-1995	JP 7203998	A	08-08-1995
WO 9627025	A	06-09-1996	AU 5171696	A	18-09-1996
WO 9106678	A	16-05-1991	CA 2044616	A	27-04-1991
			EP 0450060	A	09-10-1991
WO 9632504	A	17-10-1996	AU 5544696	A	30-10-1996
			EP 0830460	A	25-03-1998
WO 9636737	A	21-11-1996	AU 5797696	A	29-11-1996
DE 19515552	A	31-10-1996	WO 9634114	A	31-10-1996
			EP 0826067	A	04-03-1998

Form PCT/ISA/210 (patent family annex) (July 1992)